Knowledge-Bridged Causal Interaction Network for Causal Emotion Entailment

Weixiang Zhao, Yanyan Zhao *, Zhuojun Li, Bing Qin
Research Center for Social Computing and Information Retrieval
Harbin Institute of Technology, China
{wxzhao, yyyzhao, zjli, qinb}@ir.hit.edu.cn

Abstract
Causal Emotion Entailment aims to identify causal utterances that are responsible for the target utterance with a non-neutral emotion in conversations. Previous works are limited in thorough understanding of the conversational context and accurate reasoning of the emotion cause. To this end, we propose Knowledge-Bridged Causal Interaction Network (KBCIN) with commonsense knowledge (CSK) leveraged as three bridges. Specifically, we construct a conversational graph for each conversation and leverage the event-centered CSK as the semantics-level bridge (S-bridge) to capture the deep inter-utterance dependencies in the conversational context via the CSK-Enhanced Graph Attention module. Moreover, social-interaction CSK serves as emotion-level bridge (E-bridge) and action-level bridge (A-bridge) to connect candidate utterances with the target one, which provides explicit causal clues for the Emotional Interaction module and Action Interaction module to reason the target emotion. Experimental results show that our model achieves better performance over most baseline models. Our source code is publicly available at https://github.com/circle-hit/KBCIN.

Introduction
Emotion analysis in conversations has become an emerging topic in natural language processing (NLP) community. Most existing works mainly focus on Emotion Recognition in Conversations (ERC), which aims at predicting the emotion label for each utterance in conversations (Majumder et al. 2019; Ghosal et al. 2020; Shen et al. 2021). However, emotion-reasoning task such as recognizing the cause behind emotions in conversations is yet under-explored. More recently, Poria et.al (2021) holds that Recognizing Emotion Cause in CONversations (RECCON) is beneficial to improve interpretability and performance in affect-based models. Also, it has potential applications in several areas such as emotional support system (Peng et al. 2022) and empathetic dialog system (Gao et al. 2021). Thus, Poria et.al (2021) introduce a new task named RECCON with an annotated dataset. It includes two different sub-tasks: Causal Span Extraction (CSE) and Causal Emotion Entailment (CEE). We focus on the CEE sub-task in this paper and its goal is to predict which particular utterances in the conversation history trigger the non-neutral emotion in the target utterance.

There are two main challenges in the CEE task. First, to cope with the intermingling emotional dynamics among interlocutors, it is necessary to comprehend the deep semantic correlations between contextual utterances via effective conversational context modeling. Second, it may be difficult to accurately reason candidate utterances to the target emotion because causal clues are not always explicitly mentioned in the context but are supposed to be implied through reasoning based inference, which results in a reasoning gap between candidates and the target. However, Poria et.al (2021) simply formulate CEE as an utterance-pair classification problem, which is lack of sufficient conversational context modeling and effective emotion cause reasoning. Thus, to deal with such two challenges, we introduce commonsense knowledge (CSK) (Hwang et al. 2021) into CEE.

On the one hand, the event-centered CSK, which manifests what may happen before or after the event mentioned in an utterance, could be viewed as the semantics-level bridge (S-bridge) to connect the development of a conversation and enhance the semantic dependencies between relevant utterances, leading to the thorough understanding of the conver-
sational context. As shown in the left of Figure 1, the event in utterance #7 that PersonA wants to leave alone happens after the event PersonA tells PersonB to leave, which is associated with utterances #1 and #5.

On the other hand, according to (Moors and A. 2013), feeling and action tendency of human beings are two important components of emotion and contribute largely to offer potential causal clues for the generation of target emotions. To this end, social-interaction CSK is leveraged as the emotion-level bridge (E-bridge) and the action-level bridge (A-bridge) to connect candidate utterances with the target one according to causal clues conveyed by the feeling and action tendency from interlocutors. In Figure 1, the excuse and criticism from PersonB in utterance #4 make PersonA feel angry, which is consistent with the emotion carried by the target utterance #7. Further, the content of utterance #5 implies the action tendency of PersonA to be alone and it directly causes what she expresses in the target utterance #7.

In this paper, we propose Knowledge-Bridged Causal Interaction Network (KBCIN) to effectively carry out conversational context modeling and emotion cause reasoning. Specifically, we abstract a conversation as a conversational graph to model the inter-utterance dependencies in the conversation. Then event-centered CSKs including isAfter and isBefore are introduced and we devise CSK-Enhanced Graph Attention module to integrate CSKs as S-bridge for message passing on the graph. Further, to fill the reasoning gap between candidate utterances and the target one, social-interaction CSKs x(o)Want and x(o)React are leveraged as A-bridge and E-bridge. We design Emotional Interaction module and Actional Interaction module to accurately reason the cause of the target emotion with the help of the explicit causal clues conveyed by the two bridges. And the above three modules form the Knowledge-Bridged Causal Interaction (KBCI) block and it is paralleled as multiple heads to sufficiently model the inter-dependencies among utterances and precisely associate the target emotion with candidate utterances.

To evaluate the performance of the proposed model, we conduct extensive experiments on the benchmark dataset (Poria et al. 2021). State-of-the-art performance is achieved by us compared with the baseline models of CEE and other strong baselines on the task of Emotion Cause Extraction (ECE) and Emotion-Cause Pair Extraction (ECPE).

The main contributions of this work are summarized as follows:

- We introduce commonsense knowledge into the Causal Emotion Entailment task to fill the reasoning gap between candidate utterances and the target one.
- We propose a novel model KBCIN to perform comprehensive conversational context modeling and accurate emotion cause reasoning with commonsense knowledge as three bridges.
- Extensive experimental results over most strong baselines on the benchmark dataset demonstrate the superiority of our model.

Related Work

Causal Emotion Entailment

Poria et.al (2021) propose the task of recognizing emotion cause in conversations and define two novel sub-tasks named Causal Span Extraction (CSE) and Causal Emotion Entailment (CEE) to identify the emotion cause in the span-level and utterance-level, respectively. They formulate CEE as an utterance-pair classification task and build strong transformer-based baselines. Based on the emotion recognition model proposed by Shen et.al (2021), Li et.al (2022) construct knowledge-enhanced conversation graph and propagate causal clues through it.

Targeting at the two main challenges, we leverage CSK as three bridges to perform effective conversational context modeling and fill the reasoning gap between candidate utterances and the target to accurately identify the emotion cause.

Emotion Cause Extraction

Apart from CEE, there are other two emotion cause related tasks named Emotion Cause Extraction (ECE) and Emotion Cause Pair Extraction (ECPE).

Gui et.al (2016) construct the typical ECE dataset and the goal of this task is to extract cause clauses which express triggers leading to the target emotion expressed in the emotion clause. To effectively identify cause clauses, a popular strategy is to leverage relative position information to associate candidate clauses with the emotion clause (Ding et al. 2019; Xia, Zhang, and Ding 2019; Li et al. 2019). However, these works are proved to suffer from the problem of position bias and may not generalise well. Thus, recent works on ECE attempt to mitigate the position bias problem by enhance deeper semantic dependencies between the emotion clause and cause clauses (Yan et al. 2021).

Xia and Ding (2019) propose the ECPE task, with the goal to extract the potential pairs of emotions and corresponding causes in the document. Compared to ECE task, ECPE is more closer to practical applications. Previous works project the representations of every clause into a matrix to form candidate pairs and assign a confidence score to each pair. Ac-
cording to different ways of training, they could be cate-
gorized as two-stage method (Ding, Xia, and Yu 2020a) and
end-to-end method (Ding, Xia, and Yu 2020b; Wei, Zhao,
and Mao 2020; Bao et al. 2022).
Different from these two tasks which extracts emotion
cause in news articles, CEE focuses on the conversation sce-
nario, which is particularly challenging due to the complex
emotional interactions.

Methodology
Task Definition
First, we define the problem of the CEE task. Given
a conversation that consists of \( t \) consecutive utterances \( \{u_1, u_2, \cdots, u_t\} \) with the corresponding emotion label \( \{e_1, e_2, \cdots, e_t\} \) between two speakers, the goal of this task is to predict which particular utterances \( u_i (i \leq t) \) in the con-
versational history are responsible for the non-neutral emo-
tion \( e_i \) in the target utterance \( u_t \). And \( u_i \) is a positive example if it contains the cause of non-neutral emotion in the tar-
et utterance and a negative example otherwise. The archi-
tecture of our proposed model KBCIN is shown in Figure
3. We assume that \( \{u_1, u_2, \cdots, u_5\} \) is the input conversa-
tion and \( u_5 \) is the target utterance. Utterances from the same
speaker are in the same color.

Feature Extraction
Utterance-Level Feature Extraction. Transformer enc-
coder (Vaswani et al. 2017) is adopted as the utterance en-
coder to extract utterance-level features. Specifically, for
each utterance \( u_i = \{w_1, w_2, \cdots, w_L\} \), a special token
\([CLS]\) is concatenated to the beginning of the utterance.
Then we feed the sequence to the utterance encoder and the
representation after max-pooling from the last hidden layer
is obtained as the utterance-level feature of each utterance.

\[
c_i = \text{Maxpooling} (\text{Transformer} ([CLS], u_1, u_2, \cdots, u_L))
\]

(1)

where \( c_i \in \mathbb{R}^{d_m} \) and \( d_m \) is the dimension of hidden states
in utterance encoder. Then each utterance vector \( c_i \) is trans-
formed to the dimension of \( d_b \) with a linear projection.

Knowledge Acquisition. In this work, we use ATOMIC-
2020 (Hwang et al. 2021) as our commonsense knowledge
(CSK) base. It is a CSK graph covering social, physical, and
eventive aspects of everyday inferential knowledge.
To fully comprehend semantic dependencies among ut-
terances in the conversation and fill the reasoning gap be-
tween candidate utterances and the target, we leverage CSK
as three bridges, named semantics-level bridge (S-bridge),
emotion-level bridge (E-bridge) and action-level bridge (A-
bridge). To be more specific, we explore six types of
CSK representation are prepared for conversational context
modeling and emotion cause reasoning. They are denoted as
\( [\text{GEN}] \). where \( r \) is the type of CSK we select,
COMET would generate descriptions of inferential content
under the relation \( r \). And the hidden state representation
from the last layer of COMET are adopted as the CSK rep-
resentation. Through this, for each utterance \( u_i \), six pieces
of CSK representation are prepared for conversational context
modeling and emotion cause reasoning. They are denoted as
\( [\text{GEN}] \). where \( r \) is the type of CSK we select,

Paralleled Knowledge-Bridged Causal Interaction
Inspired by the notion of multi-head attention (Vaswani
et al. 2017), we propose paralleled Knowledge-Bridged Causal
Interaction block which is devised to fully comprehend the
conversational context and accurately reason the cause of
the non-neutral emotion in the target utterance. For each
block, it consists of three components: CSK-enhanced graph
attention module, emotional interaction module and actional
interaction module.

CSK-Enhanced Graph Attention Module. Instead of
formulating CEE as an utterance-pair classification problem
without explicit modeling of utterance interactions, we ab-
tract utterances in a conversation as a conversational graph
where the current utterance only connects to the past ut-
terances in the dialogue history. Through this, we make sure
the interaction of utterances to meet the nature of causality
that cause could only be reasoned from the past. And the
representation of each node is initialised by the correspond-
ing utterance-level feature \( c_i \). In addition, we calculate the
relative distance between the target utterance and candidate
utterances and utilize the relative position information to en-
rich the utterance representation. Since the emotion of each
utterance is proved to play an important role in CEE (Poria
et al. 2021), we also take it into consideration. Thus, the final
representation of each node is obtained by:

\[
h_i = c_i \oplus \text{pemb}_i \oplus \text{eemb}_i
\]

(2)
where \( \text{pemb}_i \) represents the relative position embedding be-
tween \( u_i \) and \( u_t \) and \( \text{eemb}_i \) is the emotion embedding of \( e_i \).

Based on the vanilla graph attention network (Velickovic
et al. 2018), we devise the CSK-enhanced graph attention
to propagate information on the conversational graph and
leverage event-centered CSK as S-bridge to measure the se-
matic dependencies among utterances. The graph attention
operated on the node representation to update it from the information of other neighbourhoods can be written as:

\[ \hat{h}_i = \sigma \left( \sum_{j \in N_i} \alpha_{ij} W_h h_j \right) \]  

(3)

where \( N_i \) is the neighbors of node \( i \), \( W_h \) is the trainable weight and \( \sigma \) represents the nonlinearity activation function.

The weight \( \alpha_{ij} \) is utilised to measure the relevance between the current node and its neighbours. We introduce event-centered CSK \( K^af \) and \( K^bf \) into the process:

\[ \alpha_{ij} = \frac{\exp(\mathcal{F}(h_i, h_j, K^af_j, K^bf_j))}{\sum_{j' \in N_i} \exp(\mathcal{F}(h_i, h_{j'}, K^af_{j'}, K^bf_{j'}))} \]  

(4)

Different from the attention function that purely calculates attention score between utterance representations (Velickovic et al. 2018), we leverage the event-centered CSK \( K^af \) and \( K^bf \) as the S-bridge to measure the utterance dependencies.

\[ \mathcal{F}(h_i, h_j, K^af_j, K^bf_j) = \text{LeakyReLU}(a^\top [W_h h_i \parallel W_h h_j + W_e K^af_j + W_e K^bf_j]) \]  

(5)

where \( a \in \mathbb{R}^{2d_h} \) and \( W_e \in \mathbb{R}^{d_a \times d_h} \) are both trainable weight matrices.

**Emotional Interaction Module.** After the comprehensive conversational context modeling with the help of S-bridge, we leverage two types of social-interaction CSK, \( K^{ex} \) and \( K^{ar} \) as the E-bridge to fill the reasoning gap and reason the target emotion according to the emotional causal clues. This idea is inspired by the theory that feelings are one of the most important components of human’s emotion (Moors and A. 2013). Thus, the target utterance with the corresponding emotion is more relevant to those candidates that could generate the most similar emotion or feeling as what the target holds. Further, to distinguish intra-speaker dependency and inter-speaker dependency, \( K^{ex} \) and \( K^{ar} \) serve as intra E-bridge and inter E-bridge, respectively. The emotional similarity score can be obtained by:

\[ Q^{emo} = f_q(h_t + eemb_t), \quad K^{emo} = f_k(\hat{h}_i) + f_e(K^r_i) \]

\[ s_i^{emo} = \text{softmax}(\frac{Q^{emo}(K^{emo})^\top}{\sqrt{d_h}}) \]  

(6)

where \( f_q(x); f_k(x); f_e(x) \) are all linear transformations. \( t \) is the index of the target utterance and \( i (i \leq t) \) is that of candidate utterance in the dialogue history. And \( r \) if the target utterance \( u_t \) is of the same speaker with the candidate utterance \( u_i \), otherwise \( r = ow \).

Then we utilize the emotional similarity score \( s_i^{emo} \) to weight the importance of candidate utterances and enrich them with the representation of the target utterance:

\[ h_t^{emo} = s_i^{emo} v^{emo} + s_i^{emo} Q^{emo} \]

\[ v^{emo} = f_e(K^r_i) \]  

(7)

where \( f_v(x) \) is the linear transformation.

**Actional Interaction Module.** Since action tendency is another important component to reason the aroused emotion of human beings, the other two types of social-interaction CSK \( K^{ax} \) and \( K^{aw} \) function as the A-bridge to make candidate utterances connected to the target one with the implied consistent action tendency. Also, intra A-bridge and inter A-bridge are formed. The actional similarity score is obtained:

\[ Q^{act} = f'_q(h_t), \quad K^{act} = f'_k(\hat{h}_i) + f'_e(K^r_i) \]

\[ s_i^{act} = \text{softmax}(\frac{Q^{act}(K^{act})^\top}{\sqrt{d_h}}) \]  

(8)

where \( f'_q(x); f'_k(x); f'_e(x) \) are linear transformations. \( r = aw \) if the speaker of the target utterance \( u_t \) is same as that of the candidate utterance \( u_i \), otherwise \( r = ow \).
The weighed representation after actional interaction is:

\[ h_{i}^{\text{act}} = s_{i}^{\text{act}} V^{\text{act}} + s_{i}^{\text{act}} Q^{\text{act}} \]

\[ V^{\text{act}} = f_{i}'(\hat{h}_{i}) + f_{i}'(K^{r}) \]  

(9)

where \( f_{i}'(x) \) is the linear transformation.

Finally, at the end of each knowledge-bridged causal interaction block, to synthesize results in the reasoning process, we add the conversational representation \( \hat{h}_{i} \), emotional representation \( \hat{h}_{i}^{\text{emo}} \) and actional representation \( \hat{h}_{i}^{\text{act}} \) together, and the final representation of each utterance is:

\[ \hat{h}_{i} = \hat{h}_{i} + \hat{h}_{i}^{\text{emo}} + \hat{h}_{i}^{\text{act}} \]  

(10)

**Causal Utterance Prediction**

Here, taking the concatenation of the causal representations from each paralleled KBCI head as input, we utilise a causal utterance predictor to decide whether the candidate \( u_{i} \) is the cause of the target \( u_{i} \):

\[ \hat{y}_{i} = \text{sigmoid}(\text{MLP}(\|_{n=1}^{N} h_{i,n})) \]  

(11)

where \( || \) is the concatenation operation, MLP represents the multi-layer perceptron, and \( N \) is the number of knowledge-bridged causal interaction block head.

**Experiments**

**Dataset and Evaluation Metrics**

We conduct experiments on the benchmark dataset RECCON-DD. It is collected from the popular dataset DailyDialog (Li et al. 2017) with utterance-level emotion labels and the emotion cause labels are annotated by Poria et al. (2021). We only take causes of the dialogue history into consideration and the repetitive causal pairs are removed. Statistics of the processed RECCON-DD are shown in Table 1.

Following Poria et al. (2021), we report the F1 scores of both negative and positive causal pairs and the macro F1 scores of them.

**Baselines and Comparison Models**

We compare our proposed model with baselines of CEE task. Since there are only few works on CEE, we also take baselines from other emotion cause related tasks, ECE and ECPE, into consideration.

**Methods for CEE:**

- RoBERTa-Base/Large (Poria et al. 2021) adopts the widely used pretrained language model from (Liu et al. 2019). In this setting, CEE is defined as a utterance-pair classification problem. They concatenate the special token [CLS], emotion label, target utterance, candidate utterance and the dialogue history to create the input.

- KEC (Li et al. 2022) extends the Directed Acyclic Graph networks (DAG) for ERC (Shen et al. 2021) to build Knowledge Enhanced DAG networks. They leverage social CSK to boost the performance of identifying causal utterances with neutral emotion. It is worth to mention that neutral utterances are also involved as targeted utterances, leading to more negative non-causal pairs in their processed dataset.

**Methods for ECE:**

- KAG (Yan et al. 2021) proposes a novel graph-based method to alleviate the problem of position bias by leveraging the entity-related CSK to enhance the semantic dependencies between a candidate clause and an emotion clause.

- Adapted (Turcan et al. 2021) jointly performs emotion and emotion cause recognition by combining CSK via adapted knowledge models with multi-task learning.

**Methods for ECPE:**

- ECPE-2D (Ding, Xia, and Yu 2020a) proposes an end-to-end approach for emotion-cause pair extraction. The emotion-cause pairs are represented by a 2D representation scheme and 2D transformers are devised to model the interactions of different emotion-cause pairs.

- ECPE-MLL (Ding, Xia, and Yu 2020b) proposes two joint frameworks for emotion-cause pair extraction: extraction of the cause clauses corresponding to the specified emotion clause and extraction of the emotion clauses corresponding to the specified cause clause.

- RankCP (Wei, Zhao, and Mao 2020) deals with the emotion-cause pair extraction by ranking clause pair candidates in a document and proposes a one-step neural approach to perform end-to-end extraction with inter-clause modeling.

**Implementation Details**

For utterance-level feature extraction, the dimension of hidden states in utterance encoder is 768, and the number of transformer encoder layer is 8 with 10 attention heads. Layers of emotion embedding and relative position embedding are randomly initialized and the dimension of both embedding layers are 300. Also, for all representations in the following parts of KBCIN, \( d_{b} \) is set to 300. For causal utterance prediction, dimensions of MLP is set to [300, 300, 300, 1] and the dropout rate is set to 0.07. We utilize AdamW optimizer with learning rate of 4e-5 and L2 regularization of 3e-4 to train our model. And the batch size is 8. We pick the model which works best on the valid set, and then evaluate it on the test set. All of our results are averaged on 5 runs.

**Results and Analysis**

**Overall Results**

As shown in Table 2, our proposed model achieves state-of-the-art results on REECON-DD dataset. Since results of RoBERTa-Base/Large and methods for ECPE are achieved under the same dataset scale with ours, we directly refer them from Poria et al. (2021) and we reimplement KEC and methods for ECE in the same setting with us. Benefiting from the effective conversational context modeling through
S-bridge and accurate emotion cause reasoning with E-bridge and A-bridge, KBCIN achieves state-of-the-art Pos. F1 and macro scores of 68.59 and 79.12, respectively.

For ECE and ECPE baselines, they are not comparable with our proposed KBCIN. It suggests that directly transferring the methods of context modeling and target measuring for article documents may not be suitable enough under the circumstance of conversation scenario. Also, the undesired results of KAG may be ascribed to the reason that entity-related CSK enhance the semantic dependencies among clauses to some extent, but it fails to offer valuable social-interaction causal clues to reason the target emotion.

For CEE baselines, KBCIN outperforms the strong RoBERTa-Base even RoBERTa-Large, which just concatenate the target and candidate utterances with the dialogue history. And two instructive conclusions can be derived. On the one hand, it is of great importance to explicitly perform conversational context modeling and capture inter-utterance dependencies. On the other hand, accurate target-aware emotion cause reasoning also matters and causal clues provided by commonsense knowledge sever as the bridge to fill the reasoning gap. Although CSK is also utilised in KEC, it only focuses on the emotional level to reason the target emotion and the reasoning process is restricted within a certain range by a window, which may result in the problem of position bias and weaken the ability to generalize well on other datasets, where causal utterances are not located near the target utterance. Instead of depending on the position bias to identify causal utterances, KBCIN can fully understand the conversational context to enhance the deeper inter-utterance dependencies and measures all the candidate utterances through E-bridge and A-bridge to accurately reason the cause of the target emotion.

### Ablation Study

We conduct ablation studies to verify the effectiveness of different modules proposed in our model.

#### Table 2: Comparison of our model against state-of-the-art baselines of CEE, ECE and ECPE.

<table>
<thead>
<tr>
<th>Model</th>
<th>Neg. F1</th>
<th>Pos. F1</th>
<th>macro F1</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ECE Methods</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>KAG</td>
<td>86.35</td>
<td>58.18</td>
<td>72.26</td>
</tr>
<tr>
<td>Adapted</td>
<td>88.18</td>
<td>64.53</td>
<td>76.36</td>
</tr>
<tr>
<td><strong>ECPE Methods</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ECPE-2D</td>
<td>94.96</td>
<td>55.50</td>
<td>75.23</td>
</tr>
<tr>
<td>ECPE-MLL</td>
<td>94.68</td>
<td>48.48</td>
<td>71.59</td>
</tr>
<tr>
<td>RankCP</td>
<td>97.30</td>
<td>33.00</td>
<td>65.15</td>
</tr>
<tr>
<td><strong>KBCIN (Ours)</strong></td>
<td>89.65</td>
<td>68.59</td>
<td>79.12</td>
</tr>
</tbody>
</table>

#### Table 3: Results of ablation study.

<table>
<thead>
<tr>
<th>S-bridge</th>
<th>E-bridge</th>
<th>A-bridge</th>
<th>Pos. F1</th>
<th>macro F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>68.59</td>
<td>79.12</td>
</tr>
<tr>
<td>✓</td>
<td>✓</td>
<td>✗</td>
<td>67.47</td>
<td>78.63</td>
</tr>
<tr>
<td>✓</td>
<td>✗</td>
<td>✓</td>
<td>66.33</td>
<td>77.78</td>
</tr>
<tr>
<td>✗</td>
<td>✓</td>
<td>✗</td>
<td>66.92</td>
<td>78.37</td>
</tr>
<tr>
<td>✗</td>
<td>✗</td>
<td>✓</td>
<td>57.59</td>
<td>71.81</td>
</tr>
</tbody>
</table>

#### Table 4: Results of our proposed KBCIN with different ways of using emotion information.

**Effect of Knowledge Bridge** To investigate the impact of three bridges constructed by commonsense knowledge, we remove each one of the bridge individually. First, when removing the S-bridge, the way of conversational context modeling degrades to the interactions between pure context with attention mechanism. Dropped results in the second row of Table 3 demonstrate the effectiveness of incorporating the S-bridge to enhance inter-utterance dependencies, leading to a comprehensive understanding of the conversational context. Also, by removing either the E-bridge or the A-bridge in the process of reasoning the target emotion, the performance of the model degrades to a certain extent. This suggests that both E-bridge and A-bridge play an important role in offering explicit causal clues to fill the reasoning gap between the target and candidate utterances. Further, the performance drops dramatically when we remove all three bridges, which verifies our analysis of the challenge of CEE that it requires the model to effectively capture causal clues contained in the context and accurately reason candidate utterances to the target emotion.

**Impact of Emotion Information**

To further investigate the effect of emotion information of each utterance in the dialogue history, we either remove the emotion information or replace the gold emotion labels with the labels predicted by an emotion recognition model. Results are shown in Table 4. When removing the emotion information, the clear dropped performance is demonstrated in the last row of Table 4. It verifies the effectiveness of the emotion information to directly manifest the intra and inter emotional dependencies of speakers. And the reason why we test the performance of KBCIN with the predicted labels is that emotion recognition is the prerequisite process of emotion cause extraction under the circumstance of real applications, which means such gold emotion labels of utterances in the dialogue history may not be available in a practical emotion cause extraction system. To achieve this, we utilise an advanced emotion recognition model CauAIN (Zhao, Zhao, and Lu 2022). We train the model on the train set of DailyDialog (Li et al. 2017) dataset and use it to predict the emotion labels predicted by an emotion recognition model. Replacing the gold emotion labels with the labels predicted by an emotion recognition model. When removing the emotion information or replace the gold emotion labels with the labels predicted by an emotion recognition model.
labels of each utterance in the dialogue history from the test set of RECCON-DD. And the accuracy of the predicted labels recognized by CauAIN is 76.45. As shown in the second row in Table 4, the performance of KBCIN drops to a certain degree, but it still gains an obvious improvement compared to the model without any emotion information. On the one hand, it further proves that even in the presence of false labels, emotion information of each candidate utterance is still very helpful in accurately identifying the cause of the target emotion. On the other hand, for the real application, it reminds us to jointly perform emotion recognition and emotion cause extraction in conversations, which could share the correlated emotion information between two tasks and alleviate the problem of error propagation resulting from using the emotion information in the two-stage way.

**Number of KBCI Blocks**

Since KBCI is the innovative and critical component of our model for effective conversational context modeling and accurate emotion cause reasoning, we adjust different numbers of KBCI blocks for a deeper analysis of the performance. Results are shown in Figure 4. With the increasing number of KBCI blocks from the range of 1 to 5, the model with two KBCI blocks achieves the best performance. On the one hand, the model with a single KBCI block may not comprehend deep inter-utterance dependencies contained in the conversational context and it is insufficient to effectively capture causal clues for the accurate reasoning between candidate utterances and the target. On the other hand, much more redundant causal information may be captured by the model with large numbers of block, which would weaken the performance of our model.

**Case Study**

In Figure 5, we exemplify a conversation between two speakers A and B to demonstrate that our proposed KBCIN correctly identifies the casual utterances #5 and #6 for the target utterance #6 with emotion happy. The situation is that PersonB chooses suitable scarves in the store with the help of PersonA. According to the results of Emotional Interaction and Actional Interaction to reason the cause of the target emotion, we visualize the attention weights to manifest different importance of candidate utterances. On the one hand, the target utterance #6 itself is the most attended causal utterance, where PersonA gets the exact scarf and the target emotion happy could be implied by E-bridge. On the other hand, for another causal utterance #5 that attends most by our model, PersonA recommends the heavy wool scarves for PersonB in the causal utterance #5, which meets the demand of PersonA and results in the action tendency for PersonA to buy the scarf. Also, PersonB would feel happy because of the accurate recommendation to offer a pleasant shopping experience. Here, the reasoning process is aided by both E-bridge and A-bridge to capture potential causal clues for the right prediction.

**Conclusion**

In this paper, we propose novel Knowledge-Bridged Causal Interaction Network (KBCIN) for causal emotion entailment. Commonsense knowledge (CSK) is leveraged as three bridges to carry out effective conversational context modeling and accurate emotion cause reasoning. To be more specific, we abstract the conversation as a conversation graph and leverage the event-centered CSK as the semantics-level bridge (S-bridge) to enhance the deep inter-utterance dependencies by message passing on the graph with CSK-enhanced Graph Attention module. And the social-interaction CSK serves as emotion-level bridge (E-bridge) and action-level bridge (A-bridge) to provide explicit causal clues from the perspective of human’s feeling and action tendency for Emotional Interaction module and Actional Interaction module, which fills the reasoning gap between candidate utterances and the target one. Experimental results on the benchmark dataset demonstrate the effectiveness of our proposed KBCIN.
Acknowledgments

We thank the anonymous reviewers for their insightful comments and suggestions. This work was supported by the National Key RD Program of China via grant 2021YFF0901602 and the National Natural Science Foundation of China (NSFC) via grant 62176078.

References


